DOCUMENT RESUME

ED 151 380 TM 006 798

AUTHOR Cahen, Leonard S.
TITLE The Development of a Selection Battery for Phase III-B. Beginning Teacher Evaluation Study. Technical Note Series. Technical Note III-1.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
REPORT NO BTES-TN-3-1
PUB DATE 30 Nov 76
NOTE 21p.; For related document, see TM 006 880

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Gains; *Achievement Tests; Average Students; Classroom Research; Cutting Scores; Effective Teaching; Elementary Education; Elementary School Mathematics; Elementary School Students; Grade 2; Grade 5; Item Analysis; Predictive Validity; Predictor Variables; Reading; *Sampling; *Screening Tests; *Test Construction; Testing Problems; *Test Items; Test Reliability; *Test Selection; Time Factors (Learning)

ABSTRACT

The development of the Student Selection Battery for the Beginning Teacher Evaluation Study is described. This battery, which measures reading and mathematics achievement in second and fifth grade students, will be used to select target students for the study of teaching effectiveness. Methodological considerations included: the limited amount of available testing time, the selection of items that would be appropriate predictors of achievement during later testing sessions, the effect of the testing programs on students' attitudes toward school--particularly at the beginning of grade two, and the requirement that students of "average" academic ability be selected. The development of the second grade reading battery is described. It was recommended that students scoring between the 25th and 68th percentiles on this test be selected for participation in the study. The development of the other three batteries--second grade mathematics, fifth grade reading and mathematics--was similar to the first battery described. The determination of cutting scores, and the particular difficulty in selecting the lower cutting scores, are also discussed. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151380

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: TM SP

In our judgment, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

BTES

BEGINNING TEACHER EVALUATION STUDY

Technical Note Series

TECHNICAL NOTE III-1

The Development of a Selection Battery for Phase III-C

Leonard S. Cahen

November 30, 1976

TM006 798

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT

TECHNICAL NOTE III-1

The Development of a Selection Battery for Phase III-B

by

Leonard S. Cahen

November 30, 1976

INTRODUCTION

This report describes the development of a Selection Battery for choosing target students for Phase III-B of the Beginning Teacher Evaluation Study (BTES). A second report (in press) describes the procedures actually used for selecting target students from the distribution of scores obtained when the Selection Battery was used.

Background

To understand the reasons for the development of the Selection Battery, some background information is required. It is necessary to present a short review of the research strategies for the BTES and, in particular, to focus on the importance of accurately assessing the academic status of target students.

The research to be carried out during Phase III-B requires that attention be paid to the relationship of change in reading and mathematics achievement to academic learning time (ALT), and subsequently to relationships between ALT and particular characteristics of teaching and instruction (see Proposal for Phase III-B of the Beginning Teacher Evaluation Study, submitted by the Far West Laboratory for Educational Research and Development to the California Commission for Teacher Preparation and Licensing, 1975). From research carried out during Phase III-A, BTES, 1975-76, attempts were made to develop, and/or modify tests in reading and mathematics that are reactive to instructional time. It was desirable to exclude, whenever possible, tests that were heavily associated with general intelligence because such tests were probably less likely to be influenced by variations in classroom instruction (particularly ALT). It is acknowledged that it is very difficult to separate the effects of differences in general intelligence and other extra-school influences from changes in school performances which might be attributed to ALT. Nevertheless,

the current BTES achievement test battery does reflect the considerable effort which went into developing achievement measures that should be quite sensitive to ALT, while concurrently yielding an accurate assessment of the academic status of the students. A detailed discussion of the BTES test development process is provided in Filby and Dishaw (1976).

Multi-faceted outcome measures. An issue that influenced the procedures for developing the Selection Battery was the multi-faceted nature of the academic areas that were to be measured with the achievement tests to be used in Phase III-B. The research design calls for academic assessment in reading and mathematics. Within these broad subject matter areas it was necessary to include questions ranging from factual recall to highly complex applications and problem solving. At grade 2 and grade 5, for both reading and mathematics, the question became--which subset of achievement tests could be included in the Selection Battery such that this test could efficiently predict the multi-faceted array of test scores at Fall, Winter, Spring, and Fall follow-up testing sessions? These four testing sessions are referred to in BTES reports as the "A," "B," "C," and "D" testing sessions.

Testing time. A concern with the availability of testing time also influenced the development of the Selection Battery. It was acknowledged that only a limited amount of time for test administration was available. The question raised was: How could the limited testing time available best be used? Assuming that it was possible to identify promising items or scales from the achievement battery that could serve as predictors of academic performance, then two questions arise: Should an attempt be made to sample a set of items from many predictor scales, although each type of scale that predicts overall performance would require time consuming tests directions;

or should an attempt be made to use a few complete predictor scales, thereby reducing direction time and maintaining scale reliability at a reasonable level? We chose the latter strategy.

Ethics. In addition to the above concerns, an ethical issue was involved. This concerned the administration of very difficult scales to children early in the school year. The distinct possibility existed that the testing situation might create negative attitudes and feelings about school, perhaps impairing the childrens academic self-esteem, etc. The problem of negative effect was perceived to be more real with the second grade sample of children than with the fifth grade children. This issue was given attention when the various scales were considered for the Selection Battery.

The "average" student. An important concern that influenced the development of the Selection Battery was the level of academic status of the target students. In the proposal for research during Phase III-B, the Laboratory stated that "average" achieving pupils would be selected as target students. It would be a disadvantage in this research study if the school performance in reading and mathematics of the target students was "too low." On the other hand, students who were "high achievers" might receive the top point scores on the tests administered throughout the school year. For these students it would not be difficult to detect growth in achievement. It is important to point out that the "low achieving" as well as the "high achieving" students undoubtedly improve in many of the areas of achievement that are tested. The problem lies partly in the limitations of pencil and paper tests to detect these changes. The problem also occurs because changes at one level of achievement seem to represent a different type of intellectual behavior than changes in achievement at a higher or lower level of academic

status. The BTES proposal, therefore, committed the Laboratory to selecting target students within a limited range or band of entering achievement scores on reading and mathematics tests.

Summary. The development of the Selection Battery reflected a blend of purposes and strategies which led to a set of measures that would: (1) reliably identify students in a middle range of entering achievement; (2) select students who would have the potential to display academic growth on subsequent testing occasions; (3) predict subsequent achievement over a wide range of academic tasks; (4) be economical in the time required for testing, such that under one hour is needed for administration; (5) meet the Laboratory's own personal standards of content validity and ethics in testing; and (6) avoid repetition of scales in the Selection Battery in later testing occasions. The next section of the report describes the steps that led to the development of the second grade reading Selection Battery.

Grade 2 Reading Selection Battery

During June 1976, members of the BTES staff met with an advisory panel chosen by the Commission to review reading achievement and affective measures created for Phase III-B. From this meeting a blueprint for assessing reading performance over the Phase III-B school year was developed. Many additional staff meetings were held to finalize a selection of specific scales for "A," "B," "C," and "D" test administrations. When the Phase III-B testing specifications were finalized the strategy for developing a second grade reading Selection Battery, which would meet the guidelines described earlier, was developed.

In many cases, the scales to be used in Phase III-B would be identical to those used during Phase III-A of the BTES study. In other cases, the

content remained basically unchanged, but the scales were slightly altered (i.e., one or more items were changed, added, or deleted). The availability of Phase III-A data allowed for the development of statistical "models" for identifying "average" academic status target students.[1]

The first set of the "modeling" analyses examined the dependent and independent variables available from Phase III-A testing. Table 1 shows the set of dependent and independent variables. The independent variables (potential predictor scales) in this table were chosen after a review of the entire Phase III-A achievement battery. These six tests met the criteria for a selection battery stated above. The task was to select a sub-set of scales from the six achievement measures shown under "Independent Variables" in Table 1, that would provide the best overall multivariate prediction of the two sets of dependent variables also shown in Table 1. Eight dependent variables are listed for the "B," or winter testing, with the same scales being repeated at the "C," or spring testing.

Univariate frequency distributions were examined for all variables shown in Table 1. Special attention was paid to dependent variable measures where sizeable numbers of students obtained very high (80% or more of the items passed) scale scores. All scale scores are corrected for chance guessing.

1 Recently a member of the Educational Evaluation Research, Inc. (EERI) asked why we didn't "simplify" our selection of target students by using standardized test results assumed to be available in the school districts participating in the study. This approach was considered and abandoned by the BTES staff for the following reasons: (1) the use of Phase III-A BTES data allowed an examination of the predictive relationships over a wide range of BTES outcomes deemed important and stressed in Phase III-B; (2) the kinds of achievement or aptitude test results in school files would not be uniform from district to district (i.e., different tests with unknown relationships to the BTES achievement battery are used, administrations were given at different times in the year and at different school grade levels, etc.); and (3) inherent difficulties in obtaining test results from school district files.

Table 1

Independent and Dependent Variables
for the Pyschometric Development
of the Grade 2 Reading Selection Battery

Independent Variables	Variable Code	No. of Items
Decoding Dipthongs	2R06	10
Decoding Vowel	2R07	10
Sight Words	2R10	14
Context Clues	2R13	10
Root Words	2R16	14
Comprehension with Pictures	2R35	11

Dependent Variables	Variable Code	No. of Items
Blends and Digraphs	B2R1B	9
Long Vowels	B2R4+5	22
Complex Decoding	B2R8	10
Variant Consontants	B2R3	10
Word Structure	B2R17+52	51
Synonyms	B2R22	18
Pronoun Reference	B2R25	10
Comprehension of Text	B2R27-30	39

Dependent Variables	Variable Code	No. of Items
Blends and Digraphs	C2R1B	9
Long Vowels	C2R4+5	22
Complex Decoding	C2R8	10
Variant Consonants	C2R3	10
Word Structure	C2R17+52	51
Synonyms	C2R22	18
Pronoun Reference	C2R25	10
Comprehension of Text	C2R27-30	39

Table 2 shows the "B" and "C" summary testing information on the dependent variables listed in Table 1.

Blends and digraphs (2R1B) and long vowels (2R4+5) were relatively easy scales at both "B" and "C" administrations. Pronoun Reference (2R25) and Comprehension of Text (2R27-30) were difficult tests at both "B" and "C" testings. For the latter tests, few students in the average achievement status classification would obtain 80 percent or higher scores at the "B" and "C" administrations. In addition to looking at univariate distributions we examined bi-variate distributions in which scores on each dependent variable were plotted against scores on each independent variable. All plots indicated that the relationships were generally linear. However, for the easier dependent variables, curvilinear relationships were observed in the upper ranges of the independent and dependent variables. The N's for these analyses ranged from 100 to 112.

Table 3 shows the intercorrelations among the independent variables. Table 4 shows the predictive or validity coefficients between each of the six independent variables and the 16 dependent variables. The multiple correlation coefficient for each of the dependent variables predicted from the complete set of the six predictor or independent variables is also reported in Table 4, in the column labelled "R."

At each sequential stage the BTES test development team examined output and conferred on the next stages of analysis. Scales 2R7, 2R10, and 2R35 (decoding vowels, sight words, and comprehension of pictures) were selected as our best "core" set of independent variables for the Selection Battery.

The next step was to run a series of step-wise regression analyses to predict the 16 dependent variables. Scales 2R7, 2R10, and 2R35 were forced in as a "core" set. The remaining three independent variables were then

Table 2

Summary Information of Dependent Variables
at the Winter (B) and Spring (C) Test Administrations

Dependent Variable	Description	No. of Items	$\bar{X}$	Percent of Students Passing 80% or more of Test Items
B 2R1B	Blends & Digraphs	9	6.8	60.1
C 2R1B	Blends & Digraphs	9	7.5	74.1
B 2R4+5	Long Vowels	22	14.0	43.9
C 2R4+5	Long Vowels	22	16.3	53.6
B 2R8	Complex Decoding	10	4.1	17.9
C 2R8	Complex Decoding	10	6.1	42.6
B 2R3	Variant Consonants	10	4.9	35.3
C 2R3	Variant Consonants	10	6.6	56.2
B 2R17+52	Word Structure	51	17.0	9.9
C 2R17+52	Word Structure	51	27.8	30.1
B 2R22	Synonyms	18	5.8	13.1
C 2R22	Synonyms	18	9.1	24.1
B 2R25	Pronoun Reference	10	2.4	7.8
C 2R25	Pronoun Reference	10	3.9	22.3
B 2R27-30	Comprehension of Text	39	10.9	5.6
C 2R27-30	Comprehension of Text	39	17.1	15.6

Table 3

Intercorrelation of Predictor Variables
for the Creation of the Grade 2 Reading Selection Battery
(Decimals omitted)

	2R07	2R10	2R13	2R16	2R35
2R06	75	65	65	65	57
2R07		70	71	74	63
2R10			71	62	65
2R13				64	74
2R16					59

Table 4

Validity Coefficients and Multiple Coefficients of Correlation for the Prediction of the Dependent Variables from the Independent Variables Displayed in Table 3 (Decimals omitted)

Dependent Variable	2R06	2R07	2R10	2R13	2R16	2R35	R
B2R1B	46	51	46	48	37	43	56
B2R4+5	76	74	72	71	60	61	84
B2R8	63	68	66	70	66	69	80
B2R3	66	75	65	69	59	65	80
B2R17+52	73	77	74	79	67	73	88
B2R22	64	73	68	68	65	71	81
B2R25	29	43	44	39	34	43	51
B2R27-30	57	75	68	66	66	71	83
C2R1B	60	64	73	60	50	48	77
C2R4+5	77	74	72	74	59	58	86
C2R8	59	70	61	65	54	61	75
C2R3	54	58	56	65	54	63	70
C2R17+52	73	75	71	76	63	65	85
C2R22	68	76	73	72	59	70	85
C2R25	55	59	58	62	60	66	72
C2R27-30	62	72	66	68	63	77	84

allowed into the regression equation in a sequential manner depending upon the magnitude of their partial regression weights. The results indicated that little additional predictive contribution was observed after the "core" set was entered. The standard errors of estimate when the dependent variables were predicted from the three "core" variables as separate entitites were compared to the standard error of estimate when the three "core" variables were entered as a single predictor score made up of 35 items. The total score on the 35 items was selected as the predictor score for further analyses. The results of predicting the dependent variables from the composite of (unit weight) 2R7, 2R10, and 2R35 are shown in Table 5.

The last stage of the development of the Selection Battery attended to the general properties of a 35 item Selection Battery and an investigation of the conditional distributions of the 16 dependent variables for different score levels on the 35 item total test. A total 35 item score was created for each student and entered as an additional variable in the matrix of 16 dependent variables. The total score distribution of the 35 item predictor score was examined. It appeared to have acceptable characteristics. Based on 114 cases available from Phase III-A data, the observed mean on the constructed predictor test was 16.26 with a standard deviation of 10.07. The median was 15.67, indicating a slight positive skew. Working with the reliability coefficients from the three part scores (2R7, 2R10, and 2R35), the reliability coefficient of the composite was estimated to be approximately .92 with an estimated standard error of measurement of 2.82.

Regression estimates of total predicted scores for each of the dependent variables as a function of a specific cutting point on the total selection test were then prepared. In addition to this information, sets of bi-variate plots of the prediction variable and each of the dependent variables were created. These plots summarized the actuarial relationship of outcomes on

Table 5

Multiple Coefficients of Correlation and N for Predicting a Series of Dependent Variables from 2R7, 2R10, and 2R35 for the Creation of the Grade 2 Reading Selection Battery

Dependent Variable	R	N
B 2R1B	.54	112
B 2R4+5	.78	112
B 2R8	.77	112
B 2R3	.79	102
B 2R17+52	.86	101
B 2R22	.81	107
B 2R25	.50	102
B 2R27-30	.82	100

C 2R1B	.71	108
C 2R4+5	.78	108
C 2R8	.73	108
C 2R3	.67	112
C 2R17+52	.81	106
C 2R22	.84	112
C 2R25	.70	112
C 2R27-30	.82	109

the dependent variable as a function of cutting points on the Selection Battery. It was apparent earlier that the relationship of predictor variables to the dependent variables were curvilinear for some of the easier dependent variables. In these cases a regression prediction would underestimate the predicted value on the dependent variables for high scores on the selection or predictor measure. The actuarial tables were examined for the scales where it was known that ceiling effects existed. Taking into account all of the bi-variate relationships between the predictor scores and the 16 dependent variables an upper score of 22.5 and a lower score of 8.5 were chosen as a reasonable selection band within which to select students for Phase III-B research. These upper and lower scores on the selection test were approximately at the 68th and 25th percentiles, respectively. At each stage an attempt was made to consider decisions in light of the personal knowledge that existed about the tests, suggestions of the Advisory Panel, and the psychometric information available about the tests. These were the "best" or most "reasonable" bets, as determined from previously obtained data. These points were recommended as first approximations or guidelines for the selection of grade 2 students on the reading Selection Battery in Phase III-B.

Development of the Selection Batteries for Second Grade Mathematics and Fifth Grade Reading and Mathematics

The procedures for developing a Selection Battery for second grade mathematics, fifth grade reading and fifth grade mathematics, paralleled the procedures described above for the development of the second grade reading Selection Battery. In Phase III-A separate data pools were established within grade for mathematics and reading. Therefore, it was not possible to develop a second (or fifth) grade Selection Battery simultaneously for both reading and mathematics. The data reported here on the development of the second grade mathematics, fifth grade reading, and fifth grade mathematics were performed on separate data sets.

For second grade mathematics the contribution of six possible predictor measures were examined. Seven possible predictor measures were considered for fifth grade reading, and five were available for consideration for the fifth grade mathematics battery. Regression analyses and logical examinations led us to select three predictor variables as a "core set" for second grade mathematics, fifth grade reading and for fifth grade mathematics. Table 6 displays the names and codes of the variables and the number of items in each of the scales.

For all four quadrants (second grade reading and mathematics, and fifth grade reading and mathematics) the tests comprising the "core set" were basically power tests. However, a 24 item speeded test is included in the second grade mathematics Selection Battery. All tests included in the Selection Battery were to be administered as separately timed sections when the Selection Battery was actually used for the purpose of selecting target students for Phase III-B.

Table 7 provides some summary data on the Selection Battery for second grade mathematics, fifth grade reading, and fifth grade mathematics. The

Table 6

Scales Included in
Selection Battery for
Grade 2 Mathematics
Grade 5 Mathematics
Grade 5 Reading

Grade 2 Mathematics

Scale Name	Scale Code	No. of Items
Family of Facts	2M9	24
Place Value	2M13	15
Properties	2M15	8

Grade 5 Mathematics

Subtraction	5M2	10
Addition & Subtraction Decimals	5M10	10
Cross-referenced Computation	5M29	14

Grade 5 Reading

Long & Short Vowels	5R2	21
Context Clues Meaning	5R5	12
Comprehension (Letters, Signs, Rules)	5R32SS	10

Table 7

Summary Data on Selection Batteries
for
Grade 2 Mathematics
Grade 5 Reading
Grade 5 Mathematics

Grade & Subject	No. of Items	Upper Cutting Point (Percentile)	Lower Cutting Point (Percentile)	Estimated Alpha Coefficient	Estimated Standard Error of Measurement
Grade 2 Mathematics	47	21 (70)	6.01 (27)	.88	4.00
Grade 5 Reading	43	33 (59)	17 (20)	.93	2.95
Grade 5 Mathematics	34	19.99 (70)	9 (31)	.90	3.00

table shows the number of items and the upper and lower cutting points, including percentile ranks for each Battery. The table also shows the estimated alpha reliability coefficient and the estimated standard error of measurement.

The development of four separate Selection Batteries was a very time consuming activity. After carefully considering different options and their consequences it was concluded that there was no "quick and dirty" approach to the development of a set of tests that would play such a crucial role in the selection of "proper" target students to study during Phase III-B.

A point involving the selection of cutting points on the score distributions for the Selection Battery deserves mention. Operational guidelines were available for the selection of the upper cutting points. It was felt that what was needed was a selection procedure that would minimize the number of students passing 80 percent or more of the items on the dependent measures during Phase III-B. Working from the upper cutting point, a lower cutting point was chosen that hopefully would include, on the average, about nine students per class.

The choice of a lower cutting point, however, was much more difficult. The selection tests had to identify students who were "average." There was certainly a danger in setting this cutting point too low, where students could be selected who would display relatively little growth in achievement. Setting too low a score was deemed far more serious for fith grade students than for second grade students. Second grade students have had less total experience in schooling, and hence a low score might reflect a lack of opportunity to learn as well as low aptitude.

The cutting scores constitute first approximations for identifying students in both mathematics and reading in Phase III-B. The validity of the statistical modeling and decision-making based upon Phase III-A data

will, in part, be affected by the similarity of the population of students tested in Phase III-B to those tested during Phase III-A.